PERFECT IMAGE INC.
PRECISION SCANNING
CAD SERVICES

13615 NE 126th Place #450
Kirkland, WA 98034 USA
Website:www.pimage.com

Ribbon File Scanning

Document Rev 1.0 Dated: 09July2009

Definition: The process of capturing image data directly from the scan device and writing it to a continuous uncompressed stream of data (ribbon file) on a server in its unaltered state. The ribbon file is then processed using advanced capture and recognition techniques to recognize and enhance the data for review and output by an operator. Ribbon File capture is device independent. It can be used and is available currently for Film, Fiche and Aperture Card devices directly, and allows for image import to support legacy and other devices such as document scanners, large format scanners, book scanners, overhead planetary scanners, flatbed scanners and slide/film/photographic scanners.

Preface: Please understand that this analysis, evaluation and opinions are from a Service Bureau who's goal it is to create the highest quality imagery at the most economical price. This document is not intended to recommend one manufacturer or technique over another, as each piece of equipment and software has its application and there are archives where each of the devices may be appropriate. This document is intended to explain a bit of history in film capture and describe what the latest technology is and provide visual examples. While this document briefly touches on capture speed and quality, it leaves the larger discussion surrounding image capture for another document. To be forthcoming this vendor has owned/acquired three Mekel roll film scanners in 1999-2002, a Mekel fiche scanner in 2002, a DRS fiche scanner in 2003, an OCE aperture card scanner in 1995, a Nextscan Eclipse 300 in 2003 and two Flexscans in 2008 and 2009. We have seen a Sunrise film scanner in 1999 and evaluated the output from another vendor from a Mekel Mach V in 2009.

History: In production film scanning, there are four main film scanner manufacturers Mekel (www.mekel.com), Nextscan (www.nextscan.com), Sunrise (www.sunriseimaging.com) and Wicks and Wilson (www.wwl.co.uk). They each offer film, fiche and aperture card solutions.

Prior to 2002 the scanners were very expensive, very slow, the optics were quite poor, the software was very complex and it required an operator with very advanced training in order to capture the more difficult or poor quality source medias. During this era poor quality originals that lacked good contrast, had varied frame to frame or inter-frame tonality, had varied frame size or were developed improperly were very hard to capture in a single scan of a reel. This required a high amount of quality control and a significant amount of rescanning in order to guarantee complete data capture. Grayscale scanning was available as an option and it caused the scanning devices to slow to a crawl. Most

13615 NE 126th Place #450
Kirkland, WA 98034 USA
Website:www.pimage.com

customers could not accept grayscale because their viewers did not support it and they could not afford the storage to host it. The only transfer medium at the time was CD's or hard drives in a caddy, so all of these combined to make grayscale scanning a bit impractical.

From 2002 there was more competition in the market, the price came down, the speed got much faster, the optics were improved slightly and the cost of storage was drastically reduced (to include USB drives and DVDs as a transfer medium) Grayscale scanning was included as a standard and around 2005 the devices began to output grayscale and monochrome simultaneously without affecting performance. However the processing of the imagery still occurred on the scanning device either in hardware or software. The software was still very complex and it took a very highly trained operator to capture the imagery. The devices generally worked from a preset profile that had three distinct aspects or configurations. 1) Frame Recognition 2) Crop, Rotation, Mirror, Polarity and Skew Correction 3) Image Enhancement, to include Sharpening, Despeckling, Compression and Binarization. These profiles were generally set up for each customer and adjusted by the operator based on the first say 30 or so frames on the reel. The problem with this technique is that occasionally the frames on a reel can change tonality or contrast for a number of reasons. In addition film can contain varied frame dimensions and polarity. If heavy image processing were needed on the images, the CPU cycles needed for this could impact the throughput of the scanning device particularly if the device was attempting to capture monochrome and grayscale at the same time at high resolution and with significant image sharpening.

In late 2008 the industry adopted a strategy to separate the film capture from the film processing. This freed up the processors on the capture device to dedicate their CPU cycles to image capture and all of the CPU cycles that were previously dedicated to frame recognition, image processing and enhancement were now available to speed the rate of image capture. This allows the current scanners to run at fully rated speed, regardless of any detection, image processing or enhancement, since these processes occur on a separate client or server. This separation also has the added benefit that the advanced skills necessary to produce high quality output are moved off of the scanning device and onto the "ribbon server." This allows an operator with less technical skill to operate the capture device. The operator still needs to load the film, name the reel and verify that the unit is in focus and has the proper initial illumination. They also have to deal with any serve problems on the film, such as broken film, short leaders and trailers and any areas of the film where the film is fogged or improperly developed. From our experience as long as the unit is in proper focus and the film has proper leaders and trailers, the only time where rescanning is required is

13615 NE 126th Place #450
Kirkland, WA 98034 USA
Website:www.pimage.com

if the frames are fogged mid-span on the film. Fogging is generally a result of excess light, improper temperature or and imbalance of chemicals are introduced during the original development of the film. Since this will cause an extremely dark or light section of film, for which the scanners are not designed to dynamically adjust, the operator will need to manually adjust the illumination and rescan that portion of the film. This is considered a filming error, but can be corrected for through rescanning if the section of film is legible. Prior to ribbon file capture and review, these fogged areas, particularly if they were very small, would go un-noticed unless 100% of the frames were reviewed and compared to the original. Prior to 2008 a capture device could effectively skip over these areas and unless the operator was watching the screen, nothing would be written to file and the error would not be caught in QA.

The process of reviewing the ribbon file and correcting frames is called auditing. The person performing the auditing needs to be very technical and it is important and very beneficial if they have had experience with legacy technology as it will help them to recognize areas and particular frames on the film where there is a potential for problems. They will generally focus extra review on the frames with light stamps, signatures and very light or dark frames. A vendor with ribbon capture technology can have a single auditing technician support multiple scanners. We have performed a number of jobs with the scanner and server on site and our auditor performs remote auditing. This allows us to place a scanner and an operator anywhere to capture the data. Technical videos the ribbon scanning and auditing process can be access through our website at http://www.pimage.com/technicalvideos.htm

The following manufacturers have their own versions of ribbon file capture and processing. Each differ slightly in technique and user interface, but the basic premise is the same. Sunrise has Reel/Row Scan, Wicks an Wilson has Virtual Scan Station and Nextscan has Nextstar. Mekel, as of yet, does not offer the ribbon file data capture and processing. In addition, Mekel, in our opinion, attempts to confuse potential customers by stating that detection, processing and enhancement is better done at the device because it does not require extra auditing steps to produce the output. All of the devices from the other manufacturers can run in the pre-ribbon file mode just as Mekel proposes, but the industry is clearly moving away from this techniques as the new technology reduces risk of lost data, improves throughput and final image quality. That being said Mekel scanners do certainly have their applications in the industry and they are very competitively priced.

Problems with Legacy Technology: The biggest concern for legacy technology is missed frames and high incidence of rescanning required for poor

13615 NE 126th Place #450
Kirkland, WA 98034 USA
Website:www.pimage.com

quality film and filming errors. If the scanner is more than one year old and customer/vendor does not have an alternate method to guarantee 100% image capture, then it is recommended that at a minimum the customer retain the film so that they still have access to the data in some fashion.

It is even more important that data be captured and delivered in grayscale on a legacy device since they do not have the ability to dynamically adjust for binarization without rescanning. There are tools to re-binarize from grayscale however it can be time consuming.

Frame Detection: As we discussed earlier fogged portions of the film can be skipped by these devices. Prior to ribbon scanning, the devices look at the image stream histogram to find a peak in the histogram associated with a change in polarity on the film. Generally the frames on film (fiche, roll film, app card) are black with the balance of the film being white (or vice versa). This causes a peak in the histogram just before and just after the frame. The scanners use this information to find the edge of the frame in the linear direction. In the transverse direction the operator defines the maximum width of the frame and software techniques are used to find the edge of page again based on a polarity change in the histogram. The problems with this technique is that that are problems with the histogram in many cases, and an operator, in an attempt to minimize the number of frames that are not cropped properly could errantly restrict the length or width of the frame too small, causing a frame to be cropped or even skipped before the file is ever written to the drive.

Examples of histogram problems:

1. Engineering drawings: we have seen instances where mixed polarity engineering drawings are filmed on the same reel. The scanners have no problem identifying the frames where a white engineering drawing captured as the film background is black and a polarity change occurs. The problem is that older sepias and very old technology for engineering drawings had black drawings with white text. When these are filmed adjacent to white drawings with black text, the result is a black drawing with black film background and not peak in the histogram. The result is that the scanning device skips the drawing. The only way to deal with this is to set the scanner to scan lead edge to lead edge and turn cropping off, which means it writes all of the data to the drive without cropping. When a black drawing is filmed adjacent to a white drawing, they are both written to the same file. It is then very time consuming to review and crop all of the drawings. The same thing can happen with Photostats intermixed with white letters in document filming.

2. Large graphics and black areas within a page. This can also cause peaks in the histogram, causing the scanner to inaccurately detect a new frame within a page. There is a technique called a re-arm setting that forces the scanner to effectively not look for a new edge until it has scanned a certain width which can help prevent this error, but it is ineffective when small receipts or engineering drawings are mixed with documents.
3. Long documents: Long reports, long maps such as survey or road, bridge maps, well logs, etc do not scan well on these legacy devices as they generally have a maximum image width. Since they are scanning to memory, then writing the file once the trailing edge is detected, the memory can fill up, and this defines the maximum length of the file. If a long document is intermixed with other letter sized documents, it can cause the long document to be cropped.

Image Processing: Image processing includes mainly correction for image crop, skew, rotation, mirror and polarity. Generally these items are configured in a profile and adjusted slightly by the operator based on review of the first frames on the film. If the polarity, orientation or mirror of the frames are inconsistent on the film, it requires manual review of thumbnails or at full frame to review and correct. This does require at a minimum 100% of the frames to be reviewed in thumbnail. There are OCR technologies that can assist with this, but these can also introduce error, particularly if graphics are present in the frames. The problem is with image skew and crop. If the skew is incorrectly determined the corners can be cropped off. The crop definition is generally set as a definition for the all sides equally or each side individually in terms of how many points the software should test and how aggressively it should crop. If the frames are consistent on the film or if the operator is not too aggressive then there is a high probability that all the frames will be cropped properly. If they are inconsistent, the result will be extra large black borders around images or cropped images. The latter can only be correct by rescanning, the former can be corrected by manual review of the images and manual cropping. If data is cropped of prior to writing the file, the only solution is to rescan that frame. The production scanners are not well designed for individual frame scanning, however they do have a feature to advance to a specific frame. If the vendor retains the frame number in the filename or elsewhere it makes it a bit easier to rescan. Our experience is that it takes longer to find and re-scan five individual frames that it does to scan the entire roll a second time.

Image Enhancement: This is a process of improving the definition of the characters by sharpening, removing speckles on binary (monochrome, 1bit) images and binarizing (converting to monochrome) the imagery. If a legacy scanner is scanning in monochrome only, the only way to correct for improper

13615 NE 126th Place #450
Kirkland, WA 98034 USA
Website:www.pimage.com

binarization is to rescan the page. Binarzation of film images is particularly problematic since there are so many factors that impart low contrast or tonality to the resulting data including:

1. Low contrast originals like faded faxes, the bottom copies of multipart forms which are blue with very faint typed text from the carbon process which could include manual annotations in ink, dark blue or red forms with pencil, light date stamps and signatures including stamp that were not made with even pressure, highlighting in particular blue, green and orange highlighting, maps and drawings.
2. Second, third and fourth generation data. Consider a document that is faxed, then copied, then filmed, then the film is duplicated, then scanned. The digital version will be sixth generation. There are so many areas where manual or equipment problems could impart noise dirt or otherwise degrade the data that capturing these in monochrome can be very challenging.
3. Filming and development errors. Frames can have shadows, affects from light intrusion during filming and fog or developing errors associated with temperature and chemical problems during development.

If a vendor has the proper tools they can further enhance imagery post scanning as long as the images are stored in a un-enhanced version. For instance if a high level is set at the scanner and that causes ghosting or pixelization, the frame cannot be unsharpened to restore the original fidelity of the characters. Since ribbon files are stored in un-enhanced native data files, the frames can be enhanced in batch or reviewed and enhanced individually.

The Future of Ribbon Technology: As data storages costs continue to drop more and more customers will consider archiving their ribbon files. We currently offer ribbon file hosting and remote auditing to our customers, in fact we are the first service bureau to offer ribbon hosting commercially.

Since most of the frames on film are generally legible, particularly in grayscale, it means that effectively 95% of the frames are good in monochrome and 99%+ are good in grayscale. The majority of the effort and cost is to locate and correct that last 0.5-1% of the frames that have problems. By archiving and hosting ribbon files our customers will be assured that their data is captured and available just as the original reel was presented. If they choose to reduce the cost of capturing those reels electronically they can choose to host the ribbon files with us. We deliver and or host the monochrome and grayscale output from those ribbon files. If they recognize a frame that could benefit from image correction or

PERFECT IMAGE INC.
PRECISION SCANNING
CAD SERVICES

13615 NE 126th Place #450
Kirkland, WA 98034 USA
Website:www.pimage.com

enhancement, they can log onto our servers, perform the auditing on that frame, then output new versions of the gray and monochrome image.

Summary: Digital Film Capture is a very complex process. The increased competition in the industry has afforded some major advances in technology over the last ten years. In the last year we have finally reached a point where service bureaus can easily guarantee their customers that they have captured 100% of their customers data legibly. Previously this was very hard to claim and even harder to validate or deliver. Ribbon File technology provides a more economical, secure and confident method for digital film capture. Now the industry can concentrate on improving optics, clarity and capture speed.